图书情报工作
第64卷 第13期 2020年7月

公共图书馆业务数据体系的梳理分析研究*

■ 黄文彬 刘奕 贺博
北京大学信息管理系 北京 100871 公共文化服务大数据应用文化与旅游部重点实验室 北京 100871

摘 要：[目的/意义]由于各图书馆业务服务的内容与方式不同,业务过程中数字化程度与标准不一致,给数据资源的整合造成了困难,并导致相关工作举步维艰。本文对公共图书馆业务数据体系进行梳理与分析,期望促使公共图书馆业务逐步形成数据标准,让公共图书馆的大数据应用得以扩展。[方法/过程]利用扎根理论的方法,对公共图书馆内15位各部门工作负责人进行深度访谈,通过开放式编码、主轴编码和选择性编码等过程提炼出与公共图书馆业务数据有关的概念和范畴。[结果/结论]建构以用户服务类业务、资源建设类业务和运营维护类业务为核心的公共图书馆业务数据体系,涵盖20个主范畴和62个有关的概念。该体系揭示了公共图书馆业务数据的整体框架,为大数据分析提供业务理解的指导方向,并为公共图书馆业务数字化与大数据集成提供依据,笔者也依据该体系提供文献资源分析、用户画像、运营业务的大数据分析应用示例说明。

关键词：扎根理论 公共图书馆 公共图书馆业务数据体系 大数据分析应用
分类号：G258.2
DOI：10.13266/j.issn.0252-3116.2020.13.009

1 引言

2011年6月,美国麦肯锡全球研究院发布《大数据:下一个创新、竞争和生产力的前沿》[1]研究报告之后,"大数据"成为近年来各领域研究的关键词。研究机构Gartner定义"大数据"是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产[2]。大数据及其理念、理论、方法、技术、工具、应用和实践可统称为数据科学,数据科学处理流程依序为数据化、数据加工、数据分析、数据呈现和应用以及提供数据产品[3]。数据化的任务是对现实世界业务活动中人工或系统所记录的数据进行采集,数据加工则是通过对业务和数据的理解,随后对数据进行有目的的预处理、转换、分析、建模、应用等。大数据对社会现有的知识体系、生活方式、金融业务以及公共管理等产生深远的影响[4]。

公共图书馆应用大数据,除了面临各领域大数据应用共有的技术方面的挑战以外,还具有图情学科领域独特的规律[5]。在当前社会环境里,公共图书馆扮演着信息发现者、梳理者以及开发者的角色,确立具有自己特色的服务策略,找到自己的服务方向,构建自己的服务平台是公共图书馆的重要任务之一[6]。同时,公共图书馆在普适化服务的基础上,还应当为满足读者个性化需求而开展针对性服务,以提升公共图书馆服务效能、更好地满足大众的公共文化需求[7]。大数据时代的来临对公共图书馆的发展是机遇也是挑战。对机遇而言,利用大数据来实现公共图书馆服务任务主要体现在:新知识服务的建立和完善;为公共文化服务的拓展提供强有力的技术支撑[8]。目前,公共图书馆数据的主要来源是业务数据、资源数据和网络数据,业务数据指的是业务进度、业务状态、业务流程记录、业务成果等业务文档数据,包括资源与用户交互的数据,通过对这些数据进行分析挖掘,馆员可以及时准确地监测当前状况、发现业务问题、建立新服务模式以及执行有效的决策。同时,由于分析处理的数据对象越来越庞大复杂、深入挖掘的要求和标准越来越高,公共图书馆传统的信息服务必将面临着严峻的挑战[9]。

目前,我国各地公共图书馆馆藏的内容和提供的服务有所差异,共享与整合区域内公共图书馆各项资源将是提升公共文化服务水平与发展的方向,为了达

* 本文系文化与旅游部重点实验室资助项目"公共文化服务产品的人群需求倾向与模型构建研究"(项目编号:2018006)研究成果之一。
作者简介:黄文彬(ORCID:0000-0002-9174-5467),副教授,博士;刘奕(ORCID:0000-0002-5660-2014),本科生;贺博(ORCID:0000-0002-9174-5467),硕士研究生,通讯作者,E-mail:hb1996@pku.edu.cn。
收稿日期:2019-09-28 修回日期:2020-01-11 本文起止页码:58-66 本文责任编辑:王传清

成公平、互惠和共享的目标,需要联通各地公共图书馆数据并及时进行监测与分析。在大数据时代,公共图书馆服务模式创新对于公共图书馆事业的长期良性发展具有重要的理论意义与实践意义[10]。然而,目前公共图书馆在大数据分析的应用主要集中在资源采访与读者需求分析上[11],在某种程度上局限了公共图书馆进行业务的完善和服务的创新。同时,各公共图书馆的信息系统与业务服务有同有异,使得业务数据的数字化程度与标准不一致,造成数据资源的整合十分困难,制约了公共图书馆大数据应用持续发展[12]。根据笔者实地访谈调研的结果,各公共图书馆数据不一致的主要原因是:①组织管理与业务过程并不完全一致;②记录数据的方式、周期、类型、格式等不同;③统计分析的方法与用途不同;④信息系统的设计与架构不完全一致。若从大数据分析的视角出发,公共图书馆的数据不仅来自集成管理系统的业务过程中产生的数据,还有网络各平台(网站、自媒体、社交媒体平台)以及用户使用服务的相关数据(用户访问数据、用户检索数据、用户评价数据等),这些数据有着潜力巨大的利用价值[13]。

根据切入点的不同,可以将公共图书馆大数据研究分为3类:第一类研究是探讨大数据时代公共图书馆服务创新的必要性,基本不涉及具体大数据技术;第二类研究是提出几种大数据应用于公共图书馆可以改进的服务以及改进方式,重点分析服务改进,基本不涉及大数据分析具体步骤;第三类研究是提出一种大数据应用于公共图书馆可以改进的服务以及改进方式,重点研究大数据分析具体步骤,通常会提出大数据分析算法并进行实验。通过对国内外公共图书馆大数据研究进行对比分析,可以发现目前的中文文献大多是以图书馆学和情报学视角,从公共图书馆业务出发,以第二类研究为主,小部分是第一类研究;英文文献大多是以计算机科学视角,从大数据技术出发,以第三类研究为主,小部分是第二类研究。由此可以发现,目前的公共图书馆大数据研究都是从某一点切入进行研究的,都只讨论了公共图书馆大数据的一部分研究,不同切入点的研究结果经常出现互相不重叠的情况,甚至是互相不相关,缺乏整体逻辑。

为促进公共图书馆服务的优化与大数据应用的发展,需要根据公共图书馆业务服务过程产生的资源与用户相关数据形成较全面完整的体系,为数据采集标准与公共文化服务大数据应用研究提供一个重要依据。而当下不同公共图书馆的业务不尽相同、各类数据背后的意义也不完全一致,导致业务数据无法统一。因此,笔者以典型的中型公共图书馆业务为例,依据对公共图书馆相关部门人员的实地访谈,梳理实际业务中的业务关系与业务数据,从大数据分析的视角出发理解公共图书馆的业务内容及范围,利用扎根理论对公共图书馆业务数据体系进行探索、归纳,重新构建业务关系架构,并以小型公共图书馆作为验证对象,检查该架构是否普适。本研究目的是区分公共图书馆业务数据核心范畴及其主要的概念与内涵,以期为公共文化服务大数据研究工作提供理论指导与实践意义,最终实现将公共图书馆中共性较强的业务作为集成数据的对象,通过联网将数据集成到管理机构,利用该大数据的分析提供更好的服务与决策。

2 研究方法与资料分析

2.1 扎根理论

公共图书馆的部门设置不是完全以业务类型为依据的,不同部门产生的业务数据可能是结构相同的,如借阅部门、少儿借阅部门和报刊部门,提供的都是阅览和外借服务,产生的都是借阅数据,只是面向对象和文献类型不同。因此,无法直接通过公共图书馆的部门设置对业务数据进行探索,需要深入分析业务内容,自下而上地构建公共图书馆业务数据体系。扎根理论是一种质性的研究方法[14],强调在收集提取大量数据后,以此自下而上构建出新的理论,从而发现目标研究领域隐藏的结构,更好地将研究问题概念化[15]。笔者利用扎根理论对访谈记录进行开放式编码、主轴编码、选择性编码三级编码的分析程序,并在编码完成后进行理论饱和度验证。原始数据则是从首都图书馆获得的访谈记录,并且利用从金陵图书馆获得的访谈记录进行理论饱和度验证,以完成公共图书馆业务数据体系的构建。

2.2 半结构化访谈

为了从大数据分析视角去构建公共图书馆业务数据体系,笔者分别对首都图书馆和金陵图书馆相关部门共15人进行了半结构化访谈(每人用时至少120分钟)。访谈内容包括:业务部门的业务范畴、业务流程、服务内容、业务系统、数据需要等,其访谈提纲见表1;技术部门的部门职能、数据传递、数据监控与分析、线上服务等,其访谈提纲见表2。访谈目的是:通过对公共图书馆各业务相关部门负责人进行访谈,梳理公共图书馆现有的业务流程与服务内容,收集公共图书馆已有资源与业务相关的数据,构建公共图书馆业务架

表 1 针对业务部门人员访谈提纲

序号	内容
1	了解部门业务流程、服务对象、服务内容和目标。
2	您负责的业务流程中会产生哪些数据?
3	您在业务中您需要使用其他业务产生的数据吗?
4	您认为有什么其他部门的数据可以帮助到您负责的业务吗?

表 2 针对技术部门人员访谈提纲

序号	内容
1	了解部门具体职能。
2	业务部门与技术部门间的数据传递主要通过什么途径?
3	您所在部门会有定期的数据监管和统计分析吗?如何进行?
4	贵馆有哪些线上服务业务?具体如何进行?

构。笔者整理两馆的访谈录音并分别转为文本形式,整理最终得到两份不同公共图书馆的业务部门访谈记录,其中首都图书馆的访谈记录用于进行扎根理论的编码与体系的构建,而金陵图书馆的访谈记录则用于理论饱和度验证。

2.3 开放式编码

开放式编码是将访谈记录中能够用于编码的文字进行概念化和范畴化,并重新进行组合的过程。为了构建公共图书馆业务架构,在开放式编码的过程中,以公共图书馆业务范畴为研究对象,对公共图书馆每个业务部门访谈记录中所涉及到的业务内容进行编码,赋予每个业务内容以一定的概念,因此实际上首先进行的是概念化的过程。使用的概念主要来自访谈记录中的内容,同时一个部门可能对应涉及了多个业务概念。在完成概念化之后,公共图书馆每个业务部门访谈记录中涉及到的业务内容都拥有了对应的基本概念,共得到 421 个的概念化结果,表 3 列举了其中 124 个初识概念化结果。由于原始资料内容比较庞杂,需要根据初始概念之间的相关程度和相互关系,将这些概念进一步范畴化,在范畴化的过程中对这些业务概念做进一步的抽象和整合,最终得到了 62 个初步范畴结果,开放式编码结果如表 3 所示:

表 3 开放式编码结果

初步范畴化		初始概念化			
信息服务	情报服务	用户	信息开发	系统开发	无线覆盖
资源互通	数据共享	培训	系统开发	系统采购	区域联动
数字资源推广	信息资源推广	活动	平台建设	系统维护	文化传承
讲座筹办	会议筹办	会议	平台运维	系统建设	文化传播
展览筹办	演出筹办	讲座	平台宣传	系统监测	传统节日
创客空间	平台服务	比赛	浏览反馈	书籍加工	礼品赠送
资源加工	多媒体制作	论坛	异常监测	书籍修复	证书发放
多媒体制作	设施采购	展览	舆论监测	古籍修复	网页维护
平台开发	人员招聘	演出	新闻发布	古籍整理	网络监控
资源加工	媒体制作	文创	馆员培训	古籍保护	财务管理
数据监控	硬件存放	筹办	资源典藏	读者调研	薪酬税务
信息资源存储	信息资源修护	书画	资源编目	消防设备	设备采购
数据库运维	馆内系统运维	网站	资源采访	消防演练	馆内服务
人员考核	财务核算	月报	数据统计	消防管理	相关采购
决策咨询	数据库采购	年报	数据分析	馆员互调	服务器管理
网络服务	空间装修	保洁	数据采集	馆际互借	资源数字化
活动推广	采购	保卫	数据传递	读者信用	数据库采购
比赛筹办	数据统计分析	招聘	数据存储	视频监控	纪念品定制
个人办理	硬件修护	读者证	数据清理	场地维修	宣传册发放
资源采访	信息资源使用	头条号	分析报表	工作考核	实习工读生
云平台运维	分类编目	阅读节	数据监控	资源共建	微信公众号
调研分析	加工编辑	文化节	送书到家	资源共享	读者见面会
馆内服务	系统采购	数据库	数字资源	绩效管理	多媒体资源
服务推广	特殊分类编目	实体资源	决策支持	馆员管理	流通服务站
论坛筹办	数据传递	读者反馈	参考咨询	信息安全	自媒体平台
空间服务	硬件使用	宣传推广	学科服务	志愿服务	电子读者证
人员培训	网站运维	资源借阅	学科馆员	校园合作	临时阅览证
清洁	保卫	资源浏览	设备采购	场地出租	地方资源建设
监控	消防	资源复制	设备维护	学习空间	特色资源建设
数据加工	信息资源监控	资源推广	设备监测	机构代理	资源多媒体化
网络安全监控	财会税务	读者调研	系统采购	社区覆盖	共建共享平台

黄文彬，刘奕，贺博. 公共图书馆业务数据体系的梳理分析研究[J]. 图书情报工作,2020,64(13):58－66.

2.4 主轴编码

开放式编码得到的范畴之间的关系并不明确，主轴编码的主要任务是在开放式编码所形成的范畴基础上，发掘范畴之间的对应关系，进而抽象出主范畴。通过分析比较，在开放式编码形成的 62 个初始业务范畴基础上，进行主轴编码得出 20 个主业务范畴，每个主业务范畴对应多个初始业务范畴，如表 4 所示：

表 4　主轴编码内容

主范畴	初步范畴
参考咨询	信息服务/情报服务/决策咨询/调研分析
资源流通	资源互通/数据共享/网络服务
宣传推广	数字资源推广/信息资源推广/活动推广/服务推广
活动筹办	讲座筹办/会议筹办/比赛筹办/论坛筹办/展览筹办/演出筹办
基础协助	个人办理/空间服务/馆内服务
特色服务	创客空间/平台服务
文献信息资源	资源采访/分类编目/资源加工/多媒体制作
数字信息资源	数据库采购/加工编辑/多媒体制作
硬件设施资源	设施采购
空间资源	空间装修
软件资源	系统采购/平台开发
人力资源	招聘
特色资源	采购/特殊分类编目/资源加工/媒体制作
数据管理	数据统计分析/数据传递/数据监控/数据加工
硬件设施管理	硬件存放/硬件修护/硬件使用
信息资源管理	信息资源存储/信息资源修护/信息资源使用/信息资源监控
软件平台管理	网站运维/数据库运维/馆内系统运维/云平台运维/网络安全监控
基础管理	清洁/保卫/监控/消防
人员管理	培训/考核
财务管理	财务核算/财会税务

2.5 选择性编码

整理得到的主业务范畴之间存在着相互影响、互相作用的关系，通过选择性编码，可以找到核心范畴，即能梳理公共图书馆业务的核心因素，从而发现业务间的关系，这一编码过程需要使核心范畴达到饱和。笔者经过归纳分析，这 20 个主业务范畴进一步抽象成 3 个核心范畴，分别是用户服务类业务、运营维护类业务、资源建设类业务（见表 5）。其中，用户服务类里主要是与用户直接或间接发生交互并产生了数据的业务，该用户不仅是传统意义上的读者，还包括使用移动端用户、馆员、以及政府、企业、其他机构等，具有服务需求的都是公共图书馆的用户对象。资源建设类业务是对公共图书馆馆内各类资源进行的建设工作，该资源是指公共图书馆相关的资源，包含文献信息资源、数字信息资源、特色资源、空间资源、人力资源、软硬件资源等，建设公共图书馆资源的工作皆属于资源建设类业务。运营维护类业务是为了保障对用户的服务需要以及资源建设类业务的工作进行的运营维护或组织构建。通过以上的分析，从大数据分析的角度理解公共图书馆的业务将可以从用户服务、资源建设、运营维护 3 个方面入手，而资源与用户之间的交互关系构成整个公共图书馆的核心，即资源建设和运营维护业务最终都是为用户服务提供支撑，因此围绕资源建设、保存、维护、使用以及用户的属性、参与、交互、行为等数据进行系统性的采集与关联可以提供大数据的分析较全面的帮助。

2.6 理论饱和度检验

在通过三级编码完成理论构建之后，需要对理论是否已经达到饱和进行检验，如果未达到饱和，则需要重新收集资料，并发掘新的概念和范畴[16]。在利用金陵图书馆访谈记录对理论进行检验之后，并没有发现新的概念和范畴，说明所构建的理论已经达到了饱和，上述分析过程是可信的。

3 业务数据体系构建与分析

通过扎根理论的研究构建基于大数据分析视角的公共图书馆业务数据体系，具体如图 1 所示，将公共图书馆数据划分为用户服务类、资源建设类、运营维护类 3 类核心业务数据，彼此间的关系与实际业务逻辑和功能相同。公共图书馆就是为用户而存在的，以支撑用户、创新服务为使命，把握住每个用户，分析每一个用户的现实需求，并为用户需求建立相适应的服务策略[17]。公共图书馆为了满足这些特定的用户服务而建设了相关资源，而这些服务与资源皆由运营维护业务提供管理工作支撑。换句话说，用户服务依照运营维护类业务的管理规则使用资源，用户在使用或参与服务的过程中会产生与资源的交互利用数据，而这些交互利用数据也会反馈给运维业务调整管理模式或资源的建设。资源建设业务除了为用户提供公共图书馆资源与服务外，也支撑了运维业务的工作，并且资源本身的内在关联可以生成知识结构相关的数据以及被用户和管理使用的特殊属性标签数据。运营维护业务是为了优化用户服务和资源利用而进行的管理工作，在此过程中产生了各种不同的管理活动数据。3 个核心业务交互过程的用户类型、服务内容、使用方式、管理模式、资源类型等形成业务场景重现数据与工作记录数据，因此这业务数据体系的结构就体现出公共图书馆大数据的价值与应用的方向。以下将进一步描述每个核心业务数据的内容细节。

表 5　选择编码结果

01. 用户服务类业务		02. 资源建设类业务		03. 运营维护类业务	
0101. 参考咨询	010101. 信息服务	0201. 文献信息资源	020101. 资源采访	0301. 数据管理	030101. 数据统计分析
	010102. 情报服务		020102. 分类编目		030102. 数据传递
	010103. 决策咨询		020103. 资源加工		030103. 数据监控
	010104. 调研分析		020104. 多媒体制作		030104. 数据加工
0102. 资源流通	010201. 资源互通	0202. 数字信息资源	020201. 数据库采购	0302. 硬件设施管理	030201. 硬件存放
	010202. 数据共享		020202. 加工编辑		030202. 硬件修护
	010203. 网络服务		020203. 多媒体制作		030203. 硬件使用
0103. 宣传推广	010301. 数字资源推广	0203. 硬件设施资源	020301. 设施采购	0303. 信息资源管理	030301. 信息资源存储
	010302. 信息资源推广	0204. 空间资源	020401. 空间装修		030302. 信息资源修护
	010303. 活动推广	0205. 软件资源	020501. 系统采购		030303. 信息资源使用
	010304. 服务推广		020502. 平台开发		030304. 信息资源监控
0104. 活动筹办	010401. 讲座筹办	0206. 人力资源	020601. 招聘	0304. 软件平台管理	030401. 网站运维
	010402. 会议筹办	0207. 特色资源	020701. 采购		030402. 数据库运维
	010403. 比赛筹办		020702. 分类编目		030403. 馆内系统运维
	010404. 论坛筹办		020703. 资源加工		030404. 云平台运维
	010405. 展览筹办		020704. 媒体制作		030405. 网络安全监控
	010406. 演出筹办			0305. 人员管理	030501. 培训
0105. 基础协助	010501. 个人办理				030502. 考核
	010502. 空间服务			0306. 财务管理	030601. 财务核算
	010503. 馆内服务				030602. 财会税务
0106. 特色服务	010601. 创客空间			0307. 基础管理	030701. 清洁
	010602. 平台服务				030702. 保卫
					030703. 监控
					030704. 消防

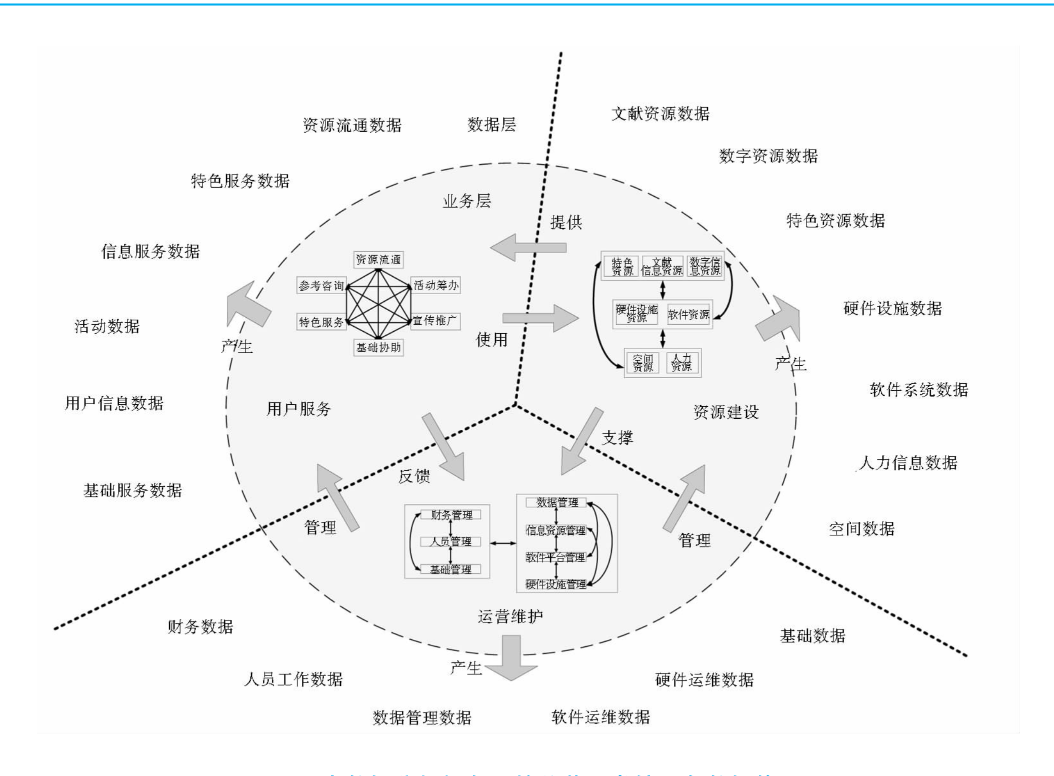


图 1　大数据分析视角下的公共图书馆业务数据体系

3.1 用户服务类业务及其数据

用户服务类业务主要包含参考咨询、资源流通、宣传推广、活动筹办、基础协助、特色服务6项主范畴的业务数据。由图1可见,这些服务业务都是可以独立运行或者彼此结合运行,各项业务都有存储数据记录的必要性,通过这些数据的结合可以挖掘出各种用户的需求、行为、偏好等。如目前信息技术不断推陈出新使得新型检索方式(如语音检索、视觉检索)逐渐为大众所熟知,管理者若想要与时俱进提升公共图书馆的服务水平,通过挖掘分析用户服务类业务的数据,依据用户类型与偏好,可以根据公共图书馆引进该技术的必要性进行分析与决策,以减少不必要的建设成本。参考咨询服务与资源流通服务的业务数据不仅包含一般用户的咨询,更记录了提供管理机构或其他单位的服务数据,有很丰富的分析价值,其分析应用将彰显出公共图书馆的能力,可以作为公共图书馆的特色服务内容之一。宣传推广、活动筹办、基础协助都是多数公共图书馆相似的业务数据,其记录方式经常不一致且数字化程度不同,使得这些数据的整合与应用成为目前的难点之一。传统的公共图书馆服务模式是以文献资源为中心,以“书本位”“刊本位”为核心理念,按照文献类型和作业内容进行业务分类的方式。然而,该服务模式忽略了用户需求,使得用户服务的业务划分与职能的切割不明确,并造成服务效率降低。有时读者所需的信息服务被分割在不同业务部门(文献阅览室、参考咨询部门、信息管理部门等),读者不能仅通过一个阅览室或者一个部门的服务就满足他的信息需求。而为了实现公共图书馆用户的个性化、专业化服务,就必须使用目前参考咨询的信息服务(表5,010101)、决策咨询(表5,010102)以及网络服务(表5,010203)的用户服务数据,再加上基础协助中个人办理(表5,010501)的用户基本属性,重组这些不同服务数据并挖掘各公共图书馆用户的习惯与特征。同时为了更好地开展用户服务类业务,需要掌握用户的信息需求,了解用户要解决的问题内容、欠缺的信息主题、文献类型以及服务形式,加强对于数字资源的宣传推广业务(表5,010301)的投入,进行有针对性的活动(表5,0104),如讲座、比赛、会议等。因此,通过对这些数据的组合与挖掘进行数据信息资源的加工(表5,020202)形成新的信息资源,并根据该分析结果以用户为中心重新设计用户服务的业务流程、完善服务内容的细节更能满足公共图书馆的用户需求并达到个性化的服务。

3.2 资源建设类业务及其数据

资源建设类业务除了文献信息资源和数字信息资源以外,还包含硬件设施资源、空间资源、软件资源、人力资源、特色资源建设等。由图1可见,公共图书馆资源的建设业务之间建立起逻辑关系先从空间资源和人力资源延伸到硬件设施资源和软件资源的基础建设,其最主要的目的是支撑文献信息资源、数字信息资源以及特色资源的建设。空间资源是公共图书馆最容易被忽略的基础资源,空间资源的监管数据可以为管理者最大化利用空间提供参考,结合空间信息、服务内容、用户数据可以优化用户服务,提升空间利用率。公共图书馆的工作性质具有专业性和服务性,与其他机构有所不同,人力资源建设过程中,公共图书馆相关工作人员的信息数据可以作为岗位匹配及招聘依据,结合运营维护类业务数据,可以更好地实现公共文化服务目标。硬件设施资源和软件资源的作用在于信息资源的收集保存、加工处理、展示应用的基础建设,同时也是提供用户直接或间接使用馆内信息服务的主要渠道之一。因此,这两种业务数据提供管理者了解目前运营上的设备使用、损坏、维修、运行等情况,对于馆内的管理与发展有一定的帮助,甚至在该馆特色发展上依据使用情况与需求提出软硬件自行开发或规格制定的模范,形成公共图书馆行业内的特殊标准。这些特殊性的工作自然就形成该馆的特色资源,因此,特色资源即可以是书籍(如古籍等)、硬件(如缩微阅读器等)、媒体(如幼儿启发、专栏视频、3D动画等),也可以是研究(如数字人文、书籍修护技术、创新文化、智库等)、系统(如共建共享平台等)等,而这些特色资源的建设本身就承载了与众不同的数据信息,通过这些数据的分析可以挖掘公共图书馆行业的新发展和新趋势,同时提高公共图书馆用户服务的创新性与趣味性。

在日益发展的信息技术支持下,公共图书馆不仅应该对用户显性的个性化需求予以重视,还应不断挖掘用户隐性的、潜在的个性化信息需求[18]。公共图书馆需要首先从采购或采访业务(表5,020101、020201)的改进做起,通过对用户信息资源需求的分析与了解,进行针对性的资源采集工作。同时在加工(表5,020103、020202)方面,对各类信息资源需要进行更深层次的揭示和组织,不仅提供文献资源描述性的特征(载体形态、题名、著者、出版者等),更需要深入到内容或语义级别的粒度,提供用户更精准知识单元的内容,并制作成新的媒体资源(表5,020104、020203),甚至可以成为馆内新的特色资源(表5,0207)。如部份

公共图书馆通过学科揭示、重新组织传播文献资料内容来提升知识服务。为了达到此目的,图书馆员需要精通某学科的专业知识并要从繁琐的文献服务中解脱出来,参与文献资料内容的揭示和读者问题的解决[19]。若通过本文的业务数据体系理解数据间的关联关系与业务关系,经由大数据分析与文本挖掘技术对信息资源依据特定需求重新组织并传播呈现,其过程将节省图书馆员的精力与时间。

3.3 运营维护类业务及其数据

运营维护类业务可以根据数据将应用的类型分成两个子类:一个子类是行政管理相关的工作,即基础管理、人员管理、财务管理,这类业务产生的数据多是日常性数据且具体可用来分析的目标较不明确,但是这些数据经过加工可以与其他数据结合进行分析。另一个子类是业务管理相关的工作,包括数据管理、硬件设施管理、信息资源管理、软件平台管理,该类业务数据记录了用户服务与公共图书馆资源之间的互动过程,通过该数据分析,可以获得许多资源的利用情形与过程,以及用户的行为偏好特征,若再与资源信息内容相结合,将获得"用户 – 服务 – 载体 – 内容"之间的时空交互关系,使得用户对信息内容的需求将展露无疑。目前,运维类业务被分割在公共图书馆内不同业务部门中,与对应的资源建设类业务和用户服务类业务被记录在不同的系统中,许多公共图书馆的运维业务过程不尽相同,因此其数据的记录格式、类型、频率等皆有所差异,使得大数据集成真正的难点工作就是在如何标准化并且有规则性的整合各公共图书馆运营维护类的业务数据。

笔者根据访谈记录,发现公共图书馆尝试的创新工作点多数都集中在运营维护类业务。如网站、公众号和其他自媒体平台是重要的宣传展示途径,对于用户吸引和公共图书馆发展有很强的支持作用,公共图书馆加强软件平台管理(表 5,0304)及硬件设施管理(表 5,0302)的工作,包括会议空间设施、加强网络接入、打印机等设备维护、馆际文献传递平台、线上服务、线上支付、新媒体体验等,用来改善有关用户体验与服务感受。随着信息技术的发展,数据统计分析、数据传递、数据监控、数据加工等数据管理业务(表 5,0301)成为目前公共图书馆大数据发展的重心,根据已有的大数据技术,只要有政策指导和资金支持,收集来自文献信息资源、数字信息资源、网站、移动平台、云平台与业务系统、硬件设施、空间等海量的数据监控,及时统计分析发现异常状况,了解馆内实时状态,预测变化趋势,甚至可以将所得的数据进行加工处理后与其他公共图书馆进行馆际合作和数据集成,形成区域性的公共图书馆大数据服务应用。

4 基于公共图书馆业务数据体系的大数据分析应用示例

当前阻碍公共图书馆大数据分析发展的因素之一是图书馆内的业务划分与业务产生的数据字段缺少统一标准,因此数据集成较为困难,各馆的业务系统无法与云平台集成系统对接。通过建立公共图书馆的业务数据体系,可以帮助公共图书馆标准化自身业务,并针对仍以印刷型资源形式存在的记录数据提供数字化标准建设的依据。从业务层面上来说,该业务数据体系能够在决策管理者规划或发布某大数据分析的执行任务时,帮助数据分析师确认该分析任务需要从哪些业务获取数据、如何获取数据以及可能用于改进某些业务等工作。换句话说,当数据分析师进行分析时,经由本文的业务数据体系得知需要数据字段 a、b、c,且分别是来自业务 A、B、C 的数据项(或是业务 A、B、C 的描述属性和评价指标等),那么该分析行为可能会对相关业务提供改进帮助,使得数据分析更加具有目的性。通过对公共图书馆业务数据的分析,一方面可以优化公共图书馆的业务流程,提高公共图书馆的工作效率;另一方面可以为公共图书馆的发展决策和战略规划提供有效的支持[20]。笔者举例说明该业务数据体系的应用,以期为公共图书馆的大数据分析提供参考。

4.1 文献资源分析

对文献资源分布状况和文献资源的利用情况进行分析,可以改进公共图书馆文献资源建设类业务,保障获取文献资源服务,应用于图书采访、调整线上资源结构等具体业务问题。通过文献资源分析获得馆内各类型资源的馆藏数量,与用户数据相结合将可获得用户偏好与习惯,并发现馆内各类型主题资源对于用户需求的满足程度。基于此,资源馆藏数量是利用资源采访业务的记录中的资源类型、采购数量这两项相关数据分析得出的,因此,这一分析结果可以作为改进资源采访业务(表 5,0201)的依据,预测某些文献在将来一段时间的使用频率进而提高用户服务的品质。如加大对于储量小、需求大的信息资源的采购力度,同时适当缩减对于储量超过用户需求的资源的购入,使得馆内资源结构更加合理,用于资源采访的资金投入更有效率;另外,资源分析得到的资源分布、资源使用、资源状

态、资源流量等,可以单独用于改进资源的加工过程,如通过资源分析得到数字化资源的种类与储量,这是使用了实体资源数字化业务记录中的资源种类与资源储量等数据,因此可以用于改进实体资源数字化业务(表5,020103),对于明显数字化程度不够的资源类型,可以考虑对其进行数字化,以网络资源的形式提供服务,增大其流通度,更好地满足用户的需求。

4.2 用户画像分析

用户画像分析的目的是将用户形象具体化,使公共图书馆提供的信息服务更贴近用户需求,达到个性化服务的目的并作为改进用户服务类业务的保障,也可用于改进活动内容、调整网站结构等具体业务问题。用户画像需要针对用户的人口属性、借阅行为、阅读行为、参与行为、资源使用行为等进行全面且系统的分析,实际上画像是分析结果的总称,各类单项的分析任务都可以有很好的应用。如若将所有用户参与活动的行为数据进行深度挖掘分析,分析过程则需要活动组织筹划业务记录中的出席人员相关数据、活动信息获取渠道、用户个人属性、用户参与深度等数据,并可藉此推知哪类的活动对于用户会具有更强的吸引力,作为活动筹办业务(表5,0104)的改进、推广、人数预测等更具有决策性和有效性的组织依据,来举办更符合用户需求、受青睐欢迎、参与度更高的活动。用户分析也可以与资源或业务分析相结合,用于改进相关业务,如将数字资源、网络服务与用户数据相结合,利用用户访问网站的资源点击、浏览、检索等日志数据,分析用户浏览公共图书馆网站内资源与行为的特征,以此帮助馆员或决策者了解用户对于网站内资源的偏好情况,改进网站的运维业务(表5,030401),发布对于用户更具有针对性的资源的相关信息,此项分析同样可以推广到其他公共图书馆宣传发布平台,如微信公众号和头条号等自媒体平台(表5,030404)。

4.3 运营业务分析

通过运营业务分析,可以有效地对公共图书馆业务执行进行监管,是公共图书馆运维类业务的重要依据。如业务分析中的一个重要项目为运营分析,需要使用到软硬件运维记录中的使用记录和异常记录数据项,以此改进软硬件运维的业务(表5,0302、0304),具体来说,通过分析可以及时得知软硬件设施的使用状况与状态(是否异常),出现问题可以进行及时有效地处理,同时,通过分析也可以知道长期以来某一设备的使用频繁程度和损坏次数,对于高使用频率、高异常频率的设备,可能是由于需求量过大,超出了设备的符合能力,公共图书馆可以考虑适当增添此类设备,提高设备采购业务的针对性(表5,030301、030501)。业务分析与对其他对象的分析相结合也可以辅助改进其他领域的业务,如在业务分析中,提取出外界舆论信息中与公共图书馆开展的活动相关的舆论评价数据进行分析,与用户画像相结合,可以得到用户对于公共图书馆所举办的活动的情感态度和倾向,并用于改进活动的宣传推广业务(表5,010303)与筹办业务(表5,0104),使得公共图书馆可以更有针对性、更有重点地开展和宣传馆内活动,吸引用户的参与。

5 结论

笔者将扎根理论融入到整个研究过程,通过对公共图书馆各部门工作负责人的访谈数据,抽象出参考咨询、资源流通、宣传推广、活动筹办、基础协助、特色服务、文献信息资源、数字信息资源、硬件设施资源、空间资源、软件资源、人力资源、特色资源建设、数据管理、硬件设施管理、信息资源管理、软件平台管理、人员管理、财务管理20个主范畴,构建以用户服务类业务、资源建设类业务和运营维护类业务为核心的公共图书馆业务数据体系。该体系揭示了公共图书馆业务数据的整体框架,为大数据分析提供业务理解的指导方向,并为公共图书馆业务内数字化需要提供建设依据,使得公共图书馆各业务数据资源能够得到较有效的整合,并对公共图书馆的数据集成有一定的帮助。笔者也根据业务数据体系举例公共图书馆大数据分析的应用参考,针对文献资源分析、用户画像、运营业务分析3类数据分析的思路以及公共图书馆业务优化与成果应用的描述,扩展公共图书馆大数据应用的思路,为公共图书馆发展战略、建设资源、优化流程以及创新服务提供一定的启发。

参考文献:

[1] MANYIKA J, CHUI M, BROWN B, et al. Big data: the next frontier for innovation, competition, and productivity [EB/OL]. [2019－03－24]. http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation.
[2] Gartner Research Institute. Big Data [EB/OL]. [2019－06－07]. https://www.gartner.com/it-glossary/big-data.
[3] 朝乐门. 信息资源管理理论的继承与创新:大数据与数据科学视角[J]. 中国图书馆学报,2019,45(2):26－42.
[4] 李言瑞,陈超,姚瑞全,等. 大数据对情报工作的影响研究[J]. 图书情报工作,2016,60(S1):179－183,58.
[5] 何胜,熊太纯,周冰,等. 高校图书馆大数据服务现实困境与应

用模式分析[J]. 图书情报工作,2015,59(22):50 –55.
[6] 陈天伦. 公共图书馆在信息经济环境中的角色与任务[J]. 图书馆研究与工作, 2016(3):20 –22,33.
[7] 张鹏. 大数据环境下公共图书馆精准服务研究[J]. 图书馆工作与研究, 2018(9): 115 –118,123.
[8] 容春琳. 公共图书馆应用大数据的策略研究[J]. 图书馆建设, 2013(7):91 –95.
[9] 韩翠峰. 大数据带给图书馆的影响与挑战[J]. 图书与情报, 2012(5):37 –40.
[10] 孙骁骁,韩海涛. 大数据时代图书馆服务创新研究[J]. 图书情报工作,2015,59(S2):28 –30.
[11] 宋亭. 公共图书馆大数据挖掘分析应用[J]. 知识经济,2019(3):83 –84.
[12] 杨雁. 公共图书馆应用大数据的理性思考[J]. 图书馆学刊, 2014,36(7):5 –8,20.
[13] 袁芳. 大数据环境下图书馆文献资源建设模式的变革[J]. 图书情报工作,2015,59(18):91 –94.
[14] GLASER B G, STRAUSS A L. The discovery of grounded theory: strategies for qualitative research[M]. London: Weidenfeld and Nicolson, 1968.
[15] 王姗姗. 高校图书馆员工招聘偏好分析——基于扎根理论的一个探究性研究[J]. 上海高校图书情报工作研究,2016(3): 55 –58.
[16] 王法硕, 王翔. 我国政府数据开放利用的影响因素与实现路径——一项基于扎根理论的质性研究[J]. 情报杂志,2016,35(7): 151 –157.
[17] 冯国权. 互联网思维下图书馆服务变革探讨[J]. 图书情报工作,2015,59(2):25 –30,16.
[18] 叶莎莎,杜杏叶. 移动图书馆用户需求理论研究[J]. 图书情报工作,2014,58(16):50 –56.
[19] 周剑. 基于知识服务的图书馆业务重组[J]. 图书馆建设, 2003(6):74 –76.
[20] 陈卫静. 智慧图书馆在大数据环境下的智慧分析[J]. 图书情报工作,2015,59(S2):49 –52.

作者贡献说明:
黄文彬:负责提出研究思路,设计研究方案,修订最终版本;
刘奕:负责收集、清理与分析数据,起草论文;
贺博:负责访谈、收集与清洗数据,修改论文。

Study on the Business Data System of the Public Library

Huang Wenbin Liu Yi He Bo
Department of Information Management, Peking University, Beijing 100871
The Key Lab of Big Data Application in Public Culture Service, Ministry of Culture and Tourism of the People's Republic of China, Beijing 100871

Abstract: [**Purpose/significance**] Due to the similarities and differences in the content and methods of the business services of various public libraries, the degree of digitization in the business process is inconsistent with the standards, which makes the integration of data resources and relevant work difficult. The purpose of this paper is to explore the business data system of public library, and hope to promote the public library business to gradually form data standards, so that the public library's big data applications can be expanded. [**Method/process**] This paper used the grounded theory method to conduct in-depth interviews with 15 department heads in public libraries, and extracted concepts and categories related to public library business through open coding, axial coding, and selective coding. [**Result/conclusion**] The public library business data system with "User Service Business", "Resource Construction Business" and "Operation and Maintenance Business" as the core is constructed, covering 20 main categories and 62 related concepts. The system reveals the overall framework of public library business data, provides guidance for understanding big business data and a basis for the digitalization of public library business. The author also provides literature resource analysis, user portrait, and operation based on the system as examples.

Keywords: grounded theory public library library business data system big data analytics application